LCMM: a R package for the estimation of latent class mixed models for Gaussian, ordinal, curvilinear longitudinal data and/or time-to-event data

Cécile Proust-Lima

Department of Biostatistics, INSERM U897, Bordeaux Segalen University

in collaboration with
Amadou Diakité and Benoit Liquet

Department of Biostatistics, INSERM U897, Bordeaux Segalen University

Rencontres R - Bordeaux - July 2nd, 2012

Context

Linear mixed model (LMM) routinely used in longitudinal studies to describe change over time of longitudinal outcomes according to covariates

Assumptions :

(i) continuous longitudinal outcome
(ii) Gaussian random-effects and errors
(iii) linearity of the relationships with the outcome
(iv) homogeneous population
(v) missing at random data

Widely implemented : `lme`, `lmer` in R ; `proc mixed` in SAS

Objective of `lcmm`

To provide a package that extends the linear mixed model estimation to :

- heterogeneous populations (relax (iv))

  → `hlme` for latent class linear mixed models (i.e. Gaussian continuous outcome)

- other types of longitudinal outcomes : ordinal, (bounded) quantitative non-Gaussian outcomes (relax (i), (ii), (iii), (iv))

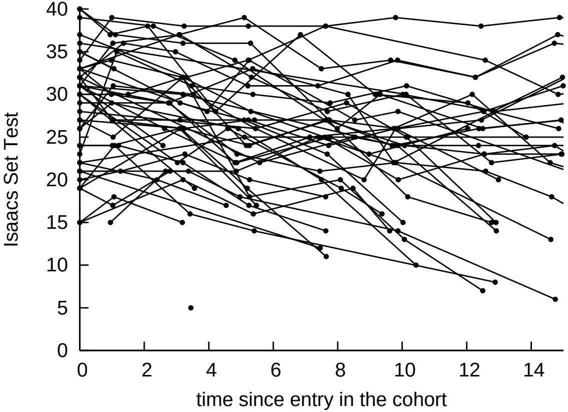
  → `lcmm` for general latent class mixed models with outcomes of different nature

- joint analysis of a time-to-event (relax (iv), (v))

  → `Jointlcmm` for joint latent class models with a longitudinal outcome and a right-censored (left-truncated) time-to-event

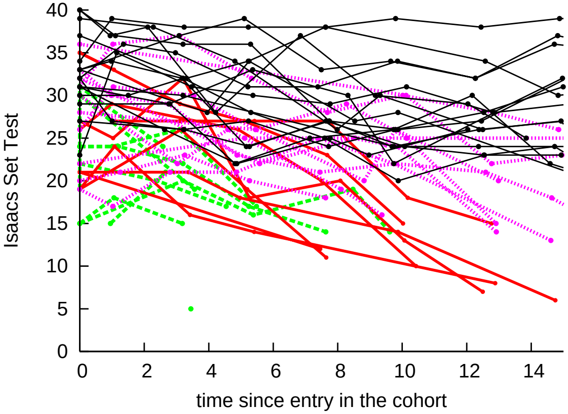
Heterogeneous profiles of trajectory

Ex : Trajectories of verbal fluency in the elderly



Heterogeneous profiles of trajectory

Ex : Trajectories of verbal fluency in the elderly



Latent class linear mixed model : the model

N subjects $(i,\ i=1,...,N)$ & G latent classes $(g,\ g=1,...,G)$

Discrete latent variable c_i for the latent group structure :
$c_i = g$ if subject i belongs to class g

$$\pi_{ig} = P(c_i = g|X_{1i}) = \frac{e^{\xi_{0g}+X_{1i}{}'\xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l}+X_{1i}{}'\xi_{1l}}}$$

Repeated measures of the longitudinal marker Y_{ij} $(j=1,...,n_i)$:

$$Y_{ij}|_{c_i=g} = Z'_{ij}u_{ig} + X'_{2ij}\beta + X'_{3ij}\gamma_g + \epsilon_{ij}$$

Z_{ij}, X_{2ij}, X_{3ij} : 3 different vectors of covariates without overlap
$u_{ig} \sim \mathcal{N}\left(\mu_g, \omega_g^2 B\right)$ & $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right)$

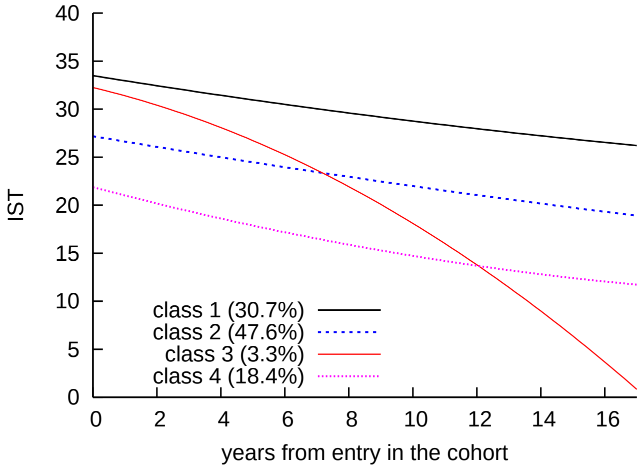
Latent class linear mixed model : the implementation

```
hlme(fixed= Y ~ Time+X1+Time:X1, random=~ Time,
subject='ID', mixture=~ Time, classmb=~ X2+X3, ng=2,
data=data_hlme, B=Binit)
```

Program details :

- Program written in Fortran90 and interfaced in R
- Marquardt algorithm for a fixed number of latent classes (strict convergence criteria : likelihood & parameters & derivatives)
- Any functions of time for the longitudinal marker
- Predictors of the trajectory (`fixed/mixture` arguments) + predictors of the class membership (`classmb`)
- Posterior functions : posterior classification, table of classification (`postprob` function), residuals, predictions (output element `pred`)

Example : profiles of verbal fluency trajectories

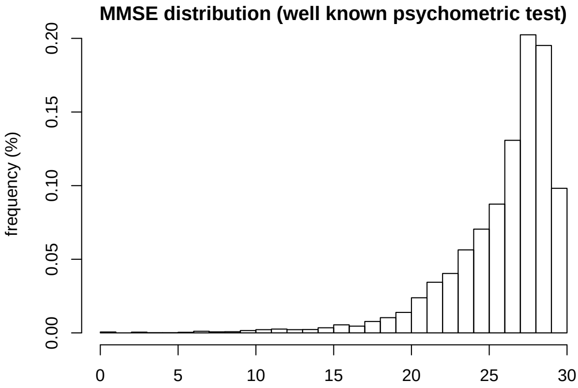


using `plot.predict` function

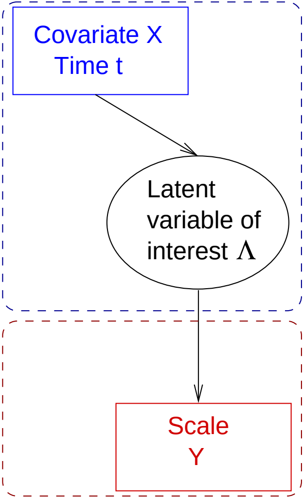
Curvilinear continuous or ordinal outcomes

Psychological & quality-of-life longitudinal scales

- bounded quantitative or ordinal longitudinal markers
- varying sensitivity to change & ceiling/floor effects ("curvilinearity")



Latent process mixed model (with latent classes)



Structural Equation
$\leftarrow$ Standard linear mixed
model or latent class linear mixed model

$$\Lambda_i(t) = \beta_0 + \beta_1 t + \beta_2 X_i + \beta_3 X_i \times t + u_{0i} + u_{1i} t$$

Equation of observation
$\leftarrow$ link between the latent process
and the outcome

$$H(Y_i(t);\eta) = \Lambda_i(t) + \epsilon_i(t)$$

with $H(\ ;\eta) =$ linear, Beta CdF, I-splines,
thresholds, etc

and $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$ - (constraints : $\beta_0 = 0$ and $\sigma = 1$)

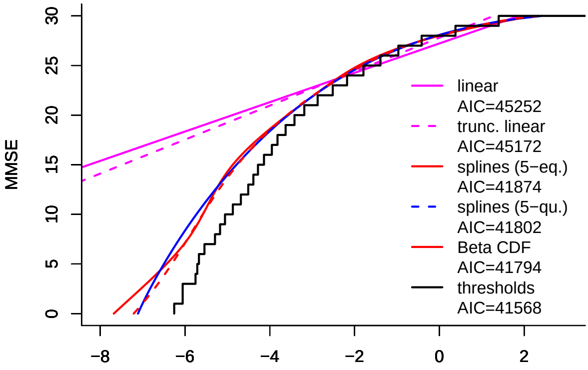
General latent class mixed model : the implementation

```
lcmm(Y ∼ Time+X1+X1:time, random=∼ Time, subject='ID',
mixture=∼ Time, classmb=∼ X2+X3, ng=2, link='splines')
```

Program details :

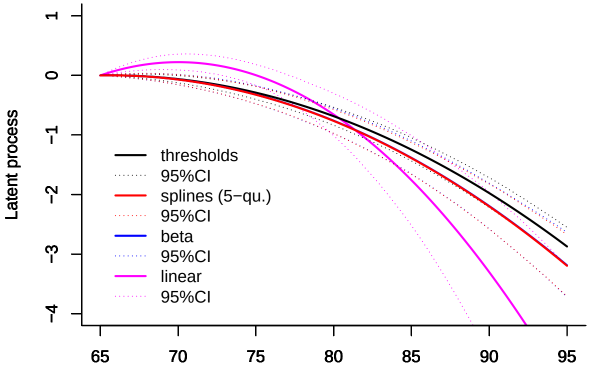
- Marquardt algorithm for a fixed number of latent classes (strict convergence criteria)
- Likelihood : closed form if continuous transformations (Jacobian decomposition) and approximated by Gauss-Hermite quadrature for threshold models (cumulated probit model)
- Goodness-of-fit : usual information criteria + “discrete” AIC (& UACV) for comparing discrete and continuous models
- Predictions in the original scale : direct or using numerical integration

Transformations of a psychometric test (MMSE))



using `plot.linkfunction` function or `estimlink` output
+ discrete AIC (and UACV)

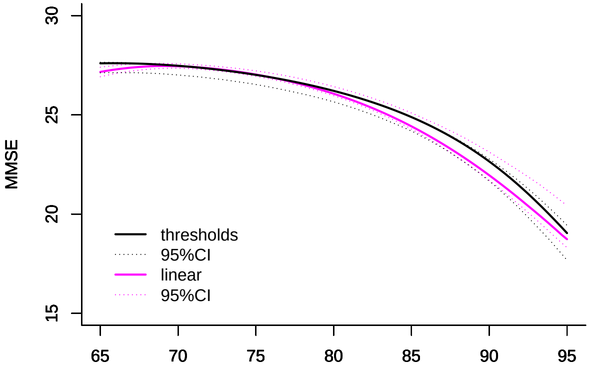
Predictions in the latent process scale



using `plot.predict` function

→ *similar technique and output with covariate effects and/or latent classes*

Predictions in the MMSE scale

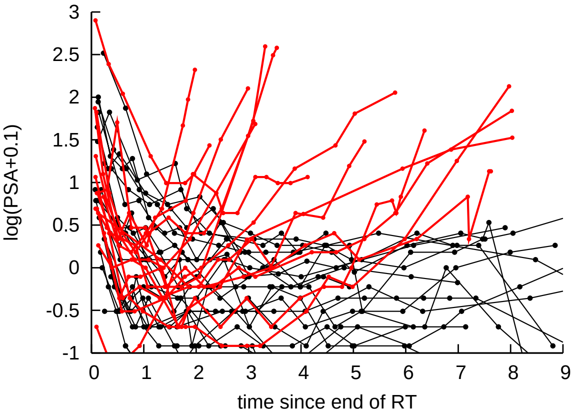


using `predictY` function with Bayesian approximation of the predictions distribution

$\rightarrow$ *similar technique and output with covariate effects and/or latent classes*

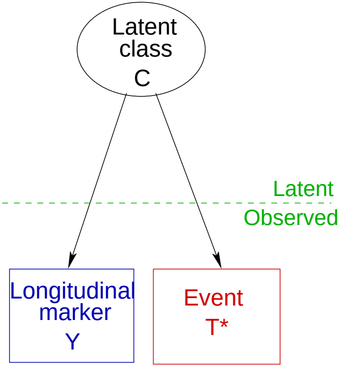
Heterogeneous profiles of trajectory in link with a clinical event

Ex : PSA trajectories after radiation therapy in Prostate Cancer



no recurrence of Prostate cancer *vs.* recurrence of Prostate cancer

Joint latent class model : the principle



Subject i (i=1...,N)
Class g (g=1,...,G)
Occasion j (j=1...,ni)

- Latent classes of subjects :

$\rightarrow$ latent class membership :

$$\pi_{ig} = P(c_i = g|X_{1i}) = \frac{e^{\xi_{0g} + X_{1i}{}^T \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + X_{1i}{}^T \xi_{1l}}}$$

- Given class g,

$\rightarrow$ specific marker evolution
(mixed model)

$\rightarrow$ specific risk of event
(prop. hazard model)

Joint latent class model : the implementation

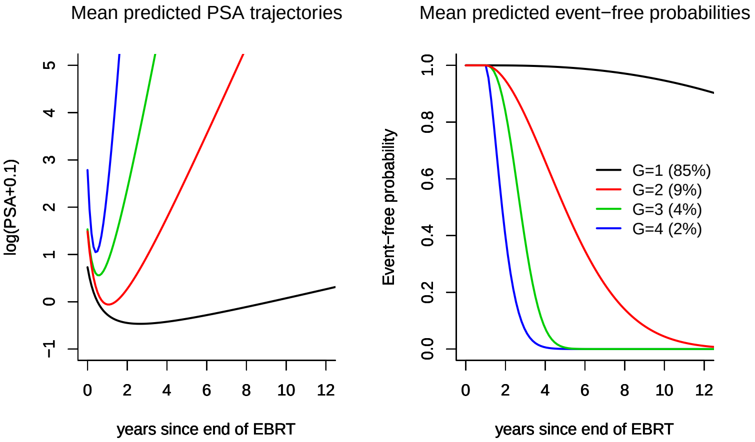
```
Jointlcmm(Y ~ Time+X1+X1:time, random=~ Time,
subject='ID', mixture=~ Time, classmb=~ X2+X3, ng=2,
survival=Surv(Tevt,Event) ~ X1 + mixture(X4),
hazard='Weibull-Specific')
```

Program details :

- Same methodology as `hlme` and `lcmm`
- PHM with different possible hazards families (splines, piecewise, weibull) and specification (common, specific, proportional over classes) + delayed entry
- Class-specific & class-common effects in the 3 submodels
- Information criteria + score test for conditional independence
- Predictive accuracy measures

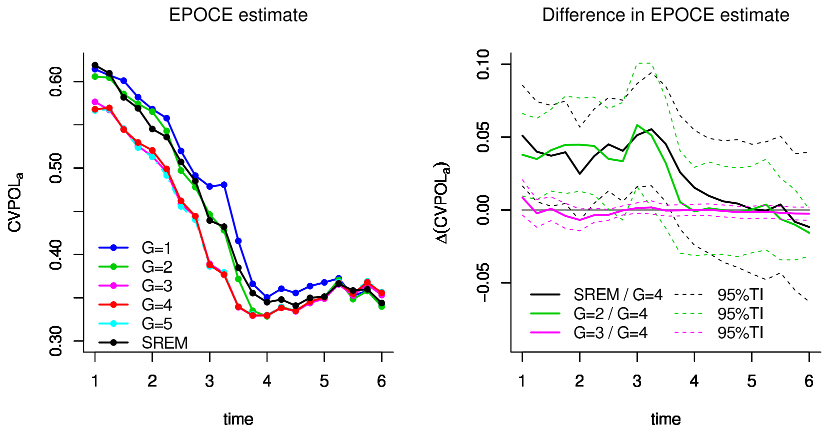
Predicted trajectories & survival functions from JLCM

(Proust-Lima, 2012)



using `plot.predict` and `plot.survival` functions

Dynamic predictive accuracy (EPOCE) (Commenges, 2012 ; Proust-Lima, 2012)



using epoce and `diffepoce` functions

Concluding remarks on lcmm R package

Extensions of the linear mixed model in MLE framework :

→ `hlme` : flexibility in covariate effects, trajectories

→ `lcmm` : analysis of subjective measures (psychological scales, patient-reported-outcomes)

→ `Jointlcmm` : computationally attractive alternative to R package JM (for shared random-effect joint models)

Planned improvements :

→ other outcomes in `Jointlcmm` (+ Poisson, logit in `lcmm`)

→ more post-fit analyses, estimation options, package spreading, tutorials ...

→ same methods for multivariate longitudinal outcomes interface in R (available in Fortran90).

Aknowledgements + references

Funding : ANR MOBIDYQ [#2010 PRSP 006 01]
+ INCa PREDYC [#2010-059]

References :

On joint latent class models and predictive accuracy :

- Commenges, Liquet & Proust-Lima (2012). *Biometrics*, early view.
- Proust-Lima, Séne, Taylor & Jacqmin-Gadda (2012). *Statistical Methods in Medical Research*, early view.
- Proust-Lima & Taylor (2009). *Biostatistics*, **10**, 535-49

On latent process mixed models :

- Proust-Lima, Amieva, Jacqmin-Gadda (2012). *paper in revision*, available on request
- Proust-Lima, Dartigues, Jacqmin-Gadda (2011). *AJE*, **174(9)**, 1077-88